SOS for Bounded Rationality

Alessio Benavoli ALESSIO@IDSIA.CH
Alessandro Facchini ALESSANDRO.FACCHINI@IDSIA.CH
Dario Piga DARIO@IDSIA.CH
Marco Zaffalon ZAFFALON@IDSIA.CH
Istituto Dalle Molle di Studi Sull'Intelligenza Artificiale (IDSIA), Lugano (Switzerland)

Abstract

In the gambling foundation of probability theory, rationality requires that a subject should always (never) find desirable all nonnegative (negative) gambles, because no matter the result of the experiment the subject never (always) decreases her money. Evaluating the nonnegativity of a gamble in infinite spaces is a difficult task. In fact, even if we restrict the gambles to be polynomials in $\mathbb{R}^n$, the problem of determining nonnegativity is NP-hard. The aim of this paper is to develop a *computable theory of desirable gambles*. Instead of requiring the subject to accept all nonnegative gambles, we only require her to accept gambles for which she can efficiently determine the nonnegativity (in particular SOS polynomials). We call this new criterion *bounded rationality*.

Keywords: bounded rationality; polynomial gambles; Sum-Of-Squares.

1. Introduction

The subjective foundation of probability by de Finetti (1937) is based on the notion of rationality (coherence or equiv. self-consistency). A subject is considered rational if she chooses her odds so that there is no bet that leads her to a sure loss (no Dutch books are possible). In this way, since odds are the inverse of probabilities, de Finetti provided a justification of Kolmogorov's axiomatisation of probability as a rationality criterion on a gambling system.[1]

Later Williams (1975) and Walley (1991) shown that it is possible to justify probability in a simpler and more elegant way. This approach is nowadays known as the *theory of desirable gambles*. To understand this gambling framework, we introduce a subject, Alice, and an experiment whose result ω belongs to a possibility space Ω (e.g., the experiment may be tossing a coin or determining the future value of a derivative instrument). When Alice is uncertain about the result ω of the experiment, we can model her beliefs about this value by asking her whether she accepts to engage in certain risky transactions, called *gambles*, whose outcome depends on the actual outcome of the experiment ω. Mathematically, a gamble is a bounded real-valued function on Ω, $g : \Omega \to \mathbb{R}$, and if Alice accepts a gamble g, this means that she commits herself to receive $g(\omega)$ utiles[2] if the experiment is performed and if the outcome of the experiment eventually happens to be the event $\omega \in \Omega$. Since $g(\omega)$ can be negative, Alice can also lose utiles and hence the desirability of a gamble depends on Alice's beliefs about Ω. Denote by $\mathcal{L}$ the set of all the gambles on Ω. Alice examines gambles in $\mathcal{L}$ and comes up with the subset $\mathcal{K}$ of the gambles that she finds desirable. How can we characterise the rationality of the assessments represented by $\mathcal{K}$?

1. De Finetti actually considered only finitely additive probabilities, while σ-additivity is assumed in Kolmogorov's axiomatisation.
2. A theoretical unit of measure of utility, for indicating a supposed quantity of satisfaction derived from an economic transaction. It is expressed in some linear utility scale.

Two obvious rationality criteria are: Alice should always accept (reject) gambles such that $g \geq 0$ ($\sup g < 0$), because no matter the result of the experiment she never (always) decreases her utiles. There is a world of difference between saying and doing. For instance, let us consider an infinite space of possibilities like $\Omega = \mathbb{R}^2$ and the gamble: $g(x_1, x_2) = 4x_1^4 + 4x_1^3 x_2 - 3x_1^2 x_2^2 + 5x_2^4$. Should Alice accept this gamble? In practice the answer to this question does not only depend on Alice's beliefs about the value of x_1 and x_2. We can in fact verify that the above polynomial can be rewritten as $(2x_1^2 - 2x_2^2 + x_1 x_2)^2 + (x_2^2 + 2x_1 x_2)^2$ and, thus, is always nonnegative. Hence, rationality implies that Alice should always accept it. However, in these cases, we must also take into account the inherent difficulty of the problem faced by Alice when she wants to determine whether a given gamble is nonnegative or not. In other words, we need to quantify the amount of computational resources needed to address rationality.

The aim of this paper is to develop a *computable theory of desirable gambles* by relaxing the two rationality criteria discussed above. In particular, instead of requiring Alice to accept all nonnegative gambles, we only require Alice to accept gambles for which she can efficiently determine the nonnegativity. We call this new criterion *bounded rationality*. The term bounded rationality was proposed by Herbert A. Simon – it is the idea that when individuals make decisions, their rationality is limited by the tractability of the decision problem, the cognitive limitations of their minds, and the time available to make the decision. Decision-makers in this view act as "satisficers", seeking a satisfactory solution rather than an optimal one. We do not propose our model as a realistic psychological model of Alice's behaviour, but we embrace the idea that the actual rationality of an agent is determined by its computational intelligence.

In this paper, we exploit the results on SOS polynomials and theory-of-moments relaxation to make numerical inferences in our theory of bounded rationality and to show that the theory of bounded rationality can be used to approximate the theory of desirable gambles. At the same time, we provide a gambling interpretation of SOS optimization. Some preliminary applications of the theoretical ideas presented in this paper can be found in Lasserre (2009); Benavoli and Piga (2016); Piga and Benavoli (2018). It is worth mentioning that a relaxation of the rationality criteria for desirability has also been investigated in Schervish et al. (2000); Pelessoni and Vicig (2016). In the first case, the work focuses on relaxations of the "avoiding sure loss" axiom, while in the second on two different criteria (additivity and positive scaling).

2. Theory of desirable gambles

In this section, we briefly introduce the theory of desirable gambles. Let us denote by $\mathcal{L}^+ = \{g \in \mathcal{L} : g \geq 0\}$ the subset of the *nonnegative gambles* and with $\mathcal{K} \subset \mathcal{L}$ the subset of the gambles that Alice finds desirable. How can we characterise the rationality of the assessments in $\mathcal{K}$?

Definition 1 *We say that $\mathcal{K}$ is a coherent set of (almost) desirable gambles (ADG) when it satisfies the following rationality criteria:*

A.1 *If* $\inf g > 0$ *then* $g \in \mathcal{K}$ *(Accepting Sure Gains);*
A.2 *If* $g \in \mathcal{K}$ *then* $\sup g \geq 0$ *(Avoiding Sure Loss);*
A.3 *If* $g \in \mathcal{K}$ *then* $\lambda g \in \mathcal{K}$ *for every* $\lambda > 0$ *(Positive Scaling);*
A.4 *If* $g, h \in \mathcal{K}$ *then* $g + h \in \mathcal{K}$ *(Additivity);*
A.5 *If* $g + \delta \in \mathcal{K}$ *for every* $\delta > 0$ *then* $g \in \mathcal{K}$ *(Closure).*

Note that A.1 and A.5 imply that $\mathcal{L}^+ \subseteq \mathcal{K}$ (including the zero gamble) (Walley, 1991; Miranda and Zaffalon, 2010). The criterion A.5 does not actually follow from rationality and can be omitted (Seidenfeld et al., 1990; Walley, 1991; Miranda and Zaffalon, 2010). However, it is useful to derive a connection between the theory of desirable gambles and probability theory and for this reason we consider it in this paper. This connection will be briefly discussed in Section 3.

To explain these rationality criteria, let us introduce a simple example: the toss of a fair coin $\Omega = \{Head, Tail\}$. A gamble g in this case has two components $g(Head) = g_1$ and $g(Tail) = g_2$. If Alice accepts g then she commits herself to receive/pay g_1 if the outcome is Heads and g_2 if Tails. Since a gamble is in this case an element of $\mathbb{R}^2$, $g = (g_1, g_2)$, we can plot the gambles Alice accepts in a 2D coordinate system with coordinate g_1 and g_2.

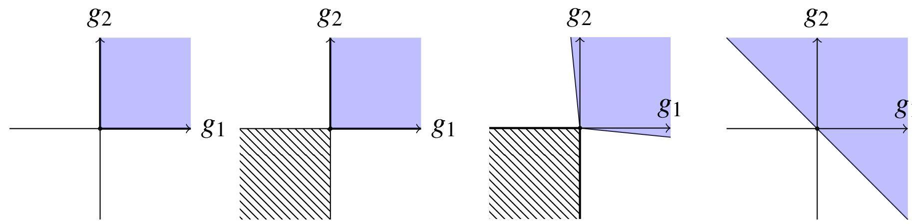
A.1 says that Alice is obviously willing to accept any gamble $g = (g_1, g_2)$ with $g_i > 0$ – Alice always accepts the first quadrant, Figure 1(a). Similarly. Alice does not accept any gamble $g = (g_1, g_2)$ with $g_i < 0$. In other words, Alice always rejects the interior of the third quadrant, Figure 1(b). This is the meaning of A.2. Then we ask Alice about $g = (-0.1, 1)$ – she loses 0.1 if Heads and wins 1 if Tails. Since Alice knows that the coin is fair, she accepts this gamble as well as all the gambles of the form νg with $\nu > 0$, because this is just a "change of currency" (this is A.3). Similarly, she accepts all the gambles $g + h$ for any $h \in \mathcal{L}^+$, since these gambles are even more favourable for her (this is basically A.4). Now, we can ask Alice about $g = (1, -0.1)$ and the argument is symmetric to the above case. We therefore obtain the following set of desirable gambles (see Figure 1(c)): $\mathcal{K}_2 = \{g \in \mathbb{R}^2 \mid 10g_1 + g_2 \geq 0 \text{ and } g_1 + 10g_2 \geq 0\}$. Finally, we can ask Alice about $g = (-1, 1)$ – she loses 1 if Heads and wins 1 if Tails. Since the coin is fair, Alice may accept or not accept this gamble. A.5 implies that she must accept it (closure). A similar conclusion can be derived for the symmetric gamble $g = (1, -1)$. Figure 1(d) is her final set of desirable gambles about the experiment concerned with the toss of a fair coin, which in a formula becomes $\mathcal{K}_3 = \{g \in \mathbb{R}^2 \mid g_1 + g_2 \geq 0\}$. Alice does not accept any other gamble. In fact, if Alice would also accept for instance $h = (-2, 0.5)$ then, since she has also accepted $g = (1.5, -1)$, i.e., $g \in \mathcal{K}_3$, she must also accept $g + h$ (because this gamble might also be favourable to her). However, $g + h = (-0.5, -0.5)$ is always negative, Alice always loses utiles in this case. In other words, by accepting $h = (-2, 0.5)$ Alice incurs a sure loss – she is irrational (A.2 is violated).

In this example, we can see that Alice's set of desirable gambles is a closed half-space, but this does not have to be the case. For instance, if Alice does not know anything about the coin, she should only accept nonnegative gambles: $\mathcal{K} = \mathcal{L}^+$. This corresponds to a state of complete ignorance, but all intermediate cases from complete belief on the probability of the coin to complete ignorance are possible. In general, $\mathcal{K}$ is a pointed (whose vertex is the origin) closed convex cone that includes $\mathcal{L}^+$ and exclude the interior of the negative orthant (this follows by A.1–A.5).

For the coin, the space of possibilities is finite and in this case Alice can check if a gamble g is nonnegative by simply examining the elements of the vector g. In this paper, we are interested in infinite spaces, in particular $\Omega = \mathbb{R}^n$, where applying the above rationality criteria is far from easy. We aim to develop a theory of *bounded rationality* for this case. Before doing that, we briefly recall the connection between ADG and probability theory.

3. Duality for ADG

Duality can be defined for general space of possibilities Ω (Walley, 1991). However, for the purpose of the present paper, we consider gambles that are bounded real-valued function on $\mathbb{R}^n$, i.e., g :



(a) (b) (c) (d)

Figure 1: Alices' sets of coherent almost desirable gambles for the experiment of tossing a fair coin.

$\mathbb{R}^n \to \mathbb{R}$. Let $\mathcal{A}$ be an algebra of subsets of $\mathbb{R}^n$ and $\mu : \mathcal{A} \to [-\infty, \infty]$ denotes a charge: that is μ a finitely additive set function of $\mathcal{A}$ (Aliprantis and Border, 2007, Ch.11). Let $\mathcal{A}_{\mathbb{R}}$ denote the algebra generated in $\mathbb{R}$ by the collection of all half open intervals (Aliprantis and Border, 2007, Th.11.8):

Theorem 2 *Every bounded $(\mathcal{A}, \mathcal{A}_{\mathbb{R}})$-measurable function is integrable w.r.t. any finite charge.*

For any $g \in \mathcal{L}$ and charge μ we can define $\int g d\mu$, that we can interpret as a linear functional $\langle \cdot, \mu \rangle$ on $\mathcal{L}$. We denote the set of all finite charges on $\mathcal{A}$ as $\mathcal{M}$ and the set of nonnegative charges as $\mathcal{M}^+$. We can then define the dual of the coherent set of desirable gambles $\mathcal{K}$ as:
$\left\{\mu \in \mathcal{M} : \int g d\mu \geq 0, \ \forall g \in \mathcal{K}\right\}$, and it can be proven that the above set is equivalent to

$$\mathcal{K}^\bullet = \left\{\mu \in \mathcal{M}^+ : \int g d\mu \geq 0, \ \forall g \in \mathcal{K}\right\}. \tag{1}$$

This follow by observing that: (i) $g = I_{\{x\}}$ (with I_x being the indicator function on $x \in \mathbb{R}^n$), is a nonnegative gamble and, therefore, is always in $\mathcal{K}$; (ii) if μ is negative for some value of $x \in \mathbb{R}^n$, i.e., $x = \tilde{x}$, then $\int I_{\tilde{x}} d\mu$ is negative too and, thus, μ cannot be in $\mathcal{K}^\bullet$. Hence, we can only focus on $\mu \in \mathcal{M}^+$. If we also impose the further requirement to $\langle \cdot, \mu \rangle$ to preserve constant gambles, in the sense that $\int c d\mu = c$, we obtain

$$\mathcal{P} = \left\{\mu \in \mathcal{M}^+ : \int g d\mu \geq 0, \ \int d\mu = 1, \ \forall g \in \mathcal{K}\right\}. \tag{2}$$

We have imposed that $\int d\mu = 1$, i.e., μ is a probability charge. Hence, it can be observed that the dual of an ADG $\mathcal{K}$ is a convex set of probability charges. The other direction of this result can be obtained by applying Hahn-Banach Theorem.

4. Finite assessments

The goal of this and next sections is to define a practical notion of desirability. To this end, we first assume that the set of gambles that Alice finds to be desirable is finitely generated. By this, we mean that there is a finite set of gambles $G = \{g_1, \ldots, g_{|G|}\}$ such that $\mathcal{K} = \mathsf{posi}(G \cup \mathcal{L}^+)$, where the posi of a set $A \subset \mathcal{L}$ is defined as $\mathsf{posi}(A) := \left\{\sum_{j=1}^{|G|} \lambda_j g_j : \ g_j \in A, \lambda_j \geq 0\right\}$, and where by $|G|$ we denote the cardinality of G. By using this definition, it is clear that whenever $\mathcal{K}$ is finitely generated, it includes all nonnegative gambles and satisfies A.3, A.4 and A.5. Once Alice has defined G and so

$\mathcal{K}$ via posi, ADG assumes that she is able to perform the following operations: to check that $\mathcal{K}$ avoids sure loss (A.2 is also satisfied); to determine the implication of desirability. It is easy to show that all above operations in ADG imply the assessment of the nonnegativity of a gamble.

Proposition 3 *Given a finite set $G \subset \mathcal{L}$ of desirable gambles, the set $\mathsf{posi}(G \cup \mathcal{L}^+)$ includes the gamble f if and only if there exist $\lambda_j \geq 0$ for $j = 1, \ldots, |G|$ such that*

$$f - \sum_{j=1}^{|G|} \lambda_j g_j \geq 0. \tag{3}$$

There are two subcases of (3) that are particularly interesting. The first is when $f = h - \lambda_0$ for some $\lambda_0 \in \mathbb{R}$ that allows us to define the concept of lower prevision Walley (1991); Miranda (2008).

Definition 4 *Assume that $\mathcal{K} = \mathsf{posi}(G \cup \mathcal{L}^+)$ is an ADG, then the solution of the following problem*

$$\sup_{\lambda_0 \in \mathbb{R}, \lambda_j \geq 0} \lambda_0, \quad s.t. \quad h - \lambda_0 - \sum_{j=1}^{|G|} \lambda_j g_j \geq 0, \tag{4}$$

is called the lower prevision of h and denoted as $\underline{P}[h]$.

From a behavioural point of view, we can reinterpret this by saying that Alice is willing to buy gamble h at price λ_0, since she is giving away λ_0 utiles while gaining h. The lower prevision is the supremum buying price for h. We can equivalently define the upper prevision of h as $\overline{P}[h] = -\underline{P}[-h]$. From Section 3, it can be easily shown that $\underline{P}[h]$ is the lower expectation of h computed w.r.t. the probability charges in $\mathcal{P}$. As a matter of fact, the dual of (4) is the moment problem: $\inf_{\mu \in \mathcal{P}} \int h d\mu$. The second subcase allows us to formulate sure loss as nonnegativity of a gamble (Walley et al., Alg.2). Let us consider $\mathcal{K} = \mathsf{posi}(G \cup \mathcal{L}^+)$ and the following problem:

$$\sup_{0 \leq \lambda_0 \leq 1, \ \lambda_j \geq 0} \lambda_0, \quad s.t. \quad -\lambda_0 - \sum_{j=1}^{|G|} \lambda_j g_j \geq 0. \tag{5}$$

$\mathcal{K}$ incurs a sure loss iff the above problem has solution $\lambda_0^* = 1$ and avoids sure loss iff $\lambda_0^* = 0$.

4.1 Complexity of inferences

When Ω is finite (e.g., coin toss), then a gamble g can also be seen as a vector in $\mathbb{R}^{|\Omega|}$, where ($|\Omega| = 2$ for the coin). Then (3) can be expressed as a linear programming problem, thus its complexity is polynomial: *Alice can check her coherence in polynomial time*. In case $\Omega = \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, solving (3) means to check the existence of real parameters $\lambda_j \geq 0$ ($j = 1, \ldots, |G|$) such that the function

$$F := f - \sum_{j=1}^{|G|} \lambda_j g_j \tag{6}$$

is non-negative in $\mathbb{R}^n$. In order to study the problem from a computational viewpoint, and avoid undecidability results, it is clear that we must impose further restrictions on the class of functions F. At the same time we would like to keep the problem general enough, in order not to lose expressiveness of the model. A good compromise can be achieved by considering the case of multivariate polynomials. The decidability of $F \geq 0$ for multivariate polynomials can be proven by means of the Tarski–Seidenberg quantifier elimination theory Tarski (1951); Seidenberg (1954).

Let $d \in \mathbb{N}$. By $\mathbb{R}_{2d}[x_1]$ we denote the set of all polynomials up to degree $2d$ in the indeterminate variable $x_1 \in \mathbb{R}$ with real-valued coefficients. With the usual definitions of addition and scalar multiplication, $\mathbb{R}_{2d}[x_1]$ becomes a vector space over the field $\mathbb{R}$ of real numbers. We can introduce a basis for $\mathbb{R}_{2d}[x_1]$ that we denote as $v_{2d}(x_1)$ where $v_j(x_1) = [1, x_1, x_1^2, \ldots, x_1^j]^\top$. We denote the dimension of $v_j(x_1)$ as $s_1(j)$ for $j = 0, 1, 2, \ldots$, e.g., $s_1(2d) = 2d+1$. Any polynomial in $\mathbb{R}_{2d}[x_1]$ can then be written as $p(x_1) = b^\top v_{2d}(x_1)$ being $b \in \mathbb{R}^{s_1(2d)}$ the vector of coefficients. We may also be interested in some subsets of $\mathbb{R}_{2d}[x_1]$ that are: (1) the subset of nonnegative polynomials that we will be denoted as $\mathbb{R}^+_{2d}[x_1]$; (2) the subset of polynomials

$$\Sigma_{2d}[x_1] = \left\{ p(x_1) \in \mathbb{R}_{2d}[x_1] \,\middle|\, p(x_1) = v_d^\top(x_1) Q v_d(x_1) \text{ with } Q \in \mathbb{R}_s^{s_1(d) \times s_1(d)},\ Q \geq 0 \right\}, \tag{7}$$

where $\mathbb{R}_s^{s_1(d) \times s_1(d)}$ is the space of $s_1(d) \times s_1(d)$ real-symmetric matrices. The polynomial $\Sigma_{2d}[x_1]$ are also called SOS polynomials, because any polynomial in $\mathbb{R}_{2d}[x_1]$ that is a sum of squares of polynomials belongs to $\Sigma_{2d}[x_1]$ and viceversa (Lasserre, 2009, Prop.2.1).

We can extend the previous framework to multivariate polynomials $\mathbb{R}_{2d}[x_1, \ldots, x_n]$, by noticing that any polynomial in $\mathbb{R}_{2d}[x_1, \ldots, x_n]$ can be written as $p(x_1, \ldots, x_n) = b^\top v_{2d}(x_1, \ldots, x_n)$ with

$$v_{2d}(x_1, \ldots, x_n) = [1, x_1, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_{n-1} x_n, x_n^2, \ldots, x_1^{2d}, \ldots, x_n^{2d}]^\top, \tag{8}$$

$b \in \mathbb{R}^{s_n(2d)}$ with $s_n(j) = \binom{n+j}{j}$ for $j = 0, 1, 2, \ldots$. Similarly to the univariate case we can define the nonnegative polynomials $\mathbb{R}^+_{2d}[x_1, \ldots, x_n]$ and the SOS polynomials $\Sigma_{2d}[x_1, \ldots, x_n]$. In the multivariate case, it is in general not true that every nonnegative polynomial is SOS or, in other words, in general $\Sigma_{2d}[x_1, \ldots, x_n] \subset \mathbb{R}^+_{2d}[x_1, \ldots, x_n]$. For instance $g(x_1, x_2) = x_1^2 x_2^2 (x_1^2 + x_2^2 - 1) + 1$ is a nonnegative polynomial that does not have a SOS representation (Lasserre, 2009, Sec.2.4). Hilbert (1888) showed the following.

Proposition 5 $\mathbb{R}^+_{2d}[x_1, \ldots, x_n] = \Sigma_{2d}[x_1, \ldots, x_n]$ *holds iff either* $n = 1$ *or* $d = 1$ *or* $(n, d) = (2, 2)$.

The problem of testing global nonnegativity of a polynomial function is in general *NP-hard*. If Alice wants to avoid the complexity associated with this problem, an alternative option is to consider a subset of polynomials for which a nonnegativity test is not *NP-hard*. The problem of testing if a given *polynomial* is SOS has polynomial complexity (we only need to check if the matrix of coefficients Q in (7) is positive-semidefinite).

5. Bounded rationality

In the bounded rationality theory we are going to represent we will work with $\Omega = \mathbb{R}^n$ and make two important assumptions. We assume that $\mathcal{L}$ is the set of multivariate polynomials of n variables and of degree less than or equal to $2d$, with $d \in \mathbb{N}$. We denote $\mathcal{L}$ as $\mathcal{L}_{2d}$ and the nonnegative polynomials as $\mathcal{L}^+_{2d}$. Note that $\mathcal{L}_{2d}$ is a vector space and A.1–A.5 are well-defined in $\mathcal{L}_{2d}$. This restriction is useful to define the computational complexity of our bounded rationality theory as a function of n and d. We now define our bounded rationality criteria, and point out the two assumptions.

Definition 6 *We say that* $C \subset \mathcal{L}_{2d}$ *is a* **bounded-rationality** *coherent set of almost desirable gambles (BADG) when it satisfies A.2–A.5 and:*

bA.1 *If* $g \in \Sigma_{2d}$ *then* $g \in C$ *(bounded accepting sure gain);*

where $\Sigma_{2d} \subset \mathcal{L}^+_{2d}$ *is the set of SOS of degree less than or equal to* $2d$.

We have seen that A.1 and A.5 imply that a coherent set of gambles must include all nonnegative gambles (and, therefore, $\mathcal{L}^+_{2d}$ that is the set of all nonnegative polynomials). Here, we restrict A.1 imposing bounded-rationality that implies that the set must only include SOS polynomials up to degree $2d$. In BADG theory, we ask Alice only to accept SOS polynomials, i.e., gambles for which she can efficiently determine the nonnegativity. Note that in Walley's terminology (Walley, 1991, Sec. 3.7.8,Appendix F) the set C is coherent relative to the vector subspace of quadratic forms $v_{2d}(x_1,\ldots,x_n)^T Q v_{2d}(x_1,\ldots,x_n)$ defined by the symmetric real matrices Q (SOS are the nonnegative gambles in this subspace, i.e., $Q \geq 0$).

In the multivariate case, we have seen that there are nonnegative polynomials that do not have a SOS representation. These polynomials should be in principle desirable for Alice in the ADG framework, but in BADG we do not enforce Alice to accept them. For this reason, BADG is a theory of bounded rationality. Note that Alice may not be able to prove that her set of desirable gambles satisfies A.2. In fact, as it has been shown in (5) this requires to check the nonnegativity of a gamble. Note however that, the requirement A.2 is weaker than A.1. In fact, while A.1 requires Alice to accept all nonnegative gambles, A.2 only requires Alice to carefully choose the gambles in G so that a sure loss is not possible. We will return on A.2 later in the section.

A BADG set C that satisfies A.2 but not A.1 can (theoretically) be turned to an ADG in $\mathcal{L}_{2d}$ by considering its extension $\mathsf{posi}(C \cup \mathcal{L}^+_{2d})$ and also to an ADG in $\mathcal{L}$ by considering its extension $\mathsf{posi}(C \cup \mathcal{L}^+)$ (note in fact that it holds $\Sigma_{2d} \subseteq \mathcal{L}^+_{2d} \subset \mathcal{L}^+$). This is important because, as it will be shown in the next sections, it will allow us to use BADG as a computable approximation of ADG.

In BADG theory, Proposition 3 is reformulated as follows.

Theorem 7 *Given a finite set* $G \subset \mathcal{L}_{2d}$ *of desirable gambles, the set* $\mathsf{posi}(G \cup \Sigma_{2d})$ *includes the gamble* f *if and only if there exist* $\lambda_j \geq 0$ *for* $j = 1,\ldots,|G|$ *such that*

$$f - \sum_{j=1}^{|G|} \lambda_j g_j \in \Sigma_{2d}. \tag{9}$$

Also in this case we can consider the gamble $f = h - \lambda_0$ for some $\lambda_0 \in \mathbb{R}$ and define the concept of lower prevision.

Definition 8 *Let* $G \subset \mathcal{L}_{2d}$ *be a finite set, and let* $C = \mathsf{posi}(G \cup \Sigma_{2d})$*. Assume that* C *is BADG, then the solution of the following problem*

$$\sup_{\lambda_0 \in \mathbb{R}, \lambda_j \geq 0} \lambda_0, \qquad s.t. \quad h - \lambda_0 - \sum_{j=1}^{|G|} \lambda_j g_j \in \Sigma_{2d}, \tag{10}$$

is called the lower prevision of h *and denoted as* $\underline{P}^*[h]$.

We can similarly use (10) to prove that $C = \mathsf{posi}(G \cup \Sigma_{2d})$ incurs a sure loss by solving the problem

$$\sup_{0 \leq \lambda_0 \leq 1,\ \lambda_j \geq 0} \lambda_0, \quad s.t. \quad -\lambda_0 - \sum_{j=1}^{|G|} \lambda_j g_j \in \Sigma_{2d}. \tag{11}$$

We have that if $\lambda_0^*=1$ then C incurs a sure loss. A similar reasoning holds for any $0\leq\lambda_0^*<1$ since, as it will be shown in Section 5.2, λ_0^* is always smaller or equal than the solution obtained in (5). This means that we cannot use (11) to prove that C avoids a sure loss. An alternative way to guarantee that $C=\mathsf{posi}(G\cup\mathcal{L}_{2d})$ avoids sure loss, is to relax (5) as

$$\lambda_0^{**}=\sup_{0\leq\lambda_0\leq 1,\ \lambda_j\geq 0}\lambda_0,\quad s.t.\quad -\lambda_0-\sum_{j=1}^{|G|}\lambda_j g_j(x_k)\geq 0,\quad k=1,\dots,M, \tag{12}$$

thus by enforcing that the constraint $-\lambda_0-\sum_{j=1}^{|G|}\lambda_j g_j\geq 0$ only holds in M (randomly generated) points $x_k\in\mathbb{R}^n$. Indeed, if $\lambda_0^{**}<1$, then the solution of problem (5) cannot be 1, thus C avoids sure loss. We will discuss this case with an example in Section 7.

5.1 Duality for BADG

We can also define the dual of a BADG. In this case, the gambles g are polynomials and the non-negative gambles that Alice accepts are SOS. Polynomials on $\mathbb{R}^n$ are not bounded functions and, therefore, we cannot use Theorem 2.[3] However, the rationality criteria A.1–A.5 do not explicitly need boundedness, but boundedness is essential to show the duality between ADG and closed convex set of probability charges, as shown in Section 3. However, since we are dealing with a vector space, we can consider its dual space $\mathcal{L}_{2d}^\bullet$, defined as the set of all linear maps $L:\mathcal{L}_{2d}\to\mathbb{R}$ (linear functionals). The dual of $C\subset\mathcal{L}_{2d}$ is defined as

$$C^\bullet=\left\{L\in\mathcal{L}_{2d}^\bullet : L(g)\geq 0,\ \forall g\in C\right\}. \tag{13}$$

Since $\mathcal{L}_{2d}$ has a basis, i.e., the monomials, if we introduce the scalars

$$y_{\alpha_1\alpha_2\dots\alpha_n}:=L(x_1^{\alpha_1}x_2^{\alpha_2},\dots,x_n^{\alpha_n})\in\mathbb{R}, \tag{14}$$

and we further assume that $y_0=L(1)=1$ (the linear functionals preserve constants), then we can rewrite $L(g)$ for any polynomial g as a function of the vector of variables $y\in\mathbb{R}^{s_n(2d)}$, whose components are the real variables $y_{\alpha_1\alpha_2\dots\alpha_n}$ defined above. This means that $\mathcal{L}_{2d}^\bullet$ is isomorphic to $\mathbb{R}^{s_n(2d)}$. We can then rewrite the dual in a simpler form. Before doing that we define the matrix $M_{n,d}(y):=L(v_d(x_1,\dots,x_n)v_d(x_1,\dots,x_n)^\top)$, where the linear operator is applied component-wise. For instance, in the case $n=1$ and $d=2$, we have that

$$M_{1,2}(y)=L(v_2(x_1)v_2(x_1)^\top)=L\left(\begin{bmatrix}1 & x_1 & x_1^2\\ x_1 & x_1^2 & x_1^3\\ x_1^2 & x_1^3 & x_1^4\end{bmatrix}\right)=\begin{bmatrix}y_0 & y_1 & y_2\\ y_1 & y_2 & y_3\\ y_2 & y_3 & y_4\end{bmatrix}.$$

We have then the following result (see for instance Lasserre, 2009).

Theorem 9 *Let C be a BADG. Then its dual is*

$$C^\bullet=\left\{y\in\mathbb{R}^{s_n(2d)} : L(g)\geq 0,\ \ M_{n,d}(y)\geq 0,\ \forall g\in C\right\}. \tag{15}$$

where $L(g)$ is completely determined by y via the definition (14).

3. For an extension of the theory of desirable gambles to unbounded gambles see Troffaes and De Cooman (2003)

Proof We have seen that any SOS in Σ_{2d} can be written as $v_d(x_1,\dots,x_n)^\top Q v_d(x_1,\dots,x_n)$ (see eq. (7)). By exploiting matrix algebra we have $v_d(x_1,\dots,x_n)^\top Q v_d(x_1,\dots,x_n)$ is equal to $Tr(Qv_d(x_1,\dots,x_n)v_d(x_1,\dots,x_n)^\top)$ with $Q \geq 0$. Now observe that because of linearity of L and trace

$$L(Tr(Qv_d(x_1,\dots,x_n)v_d(x_1,\dots,x_n)^\top)) = Tr(QL(v_d(x_1,\dots,x_n)v_d(x_1,\dots,x_n)^\top)) = Tr(QM_{n,d}(y))$$

where $M_{n,d}(y) = L(v_d(x_1,\dots,x_n)v_d(x_1,\dots,x_n)^\top$. From $L(g) \geq 0$ in (13) for any $g \in \Sigma_{2d}$ we have $Tr(QM_{n,d}(y)) \geq 0$. This means that $Tr(QM_{n,d}(y)) \geq 0 \;\; \forall Q \geq 0$. This implies that $M_{n,d}(y) \geq 0$ (it can be proven by using the eigenvalue-eigenvector decomposition of $M_{n,d}(y)$). ■

The other direction follows by Hahn-Banach Theorem. Note that when $C = \Sigma_{2d}$, its dual is

$$C^\bullet = \left\{ y \in \mathbb{R}^{s_n(2d)} : M_{n,d}(y) \geq 0 \right\}, \tag{16}$$

which corresponds to a state of ignorance: Alice only accepts nonnegative gambles.

In Section 3, by considering the space of all bounded gambles, we have showed that the dual of an ADG is a closed convex set of probability charges. In (15) there is no reference to probability. However, if the integral $\int x_1^{\alpha_1} x_2^{\alpha_2},\dots,x_n^{\alpha_n} d\mu$ is well-defined, we can interpret $y_{\alpha_1\alpha_2\dots\alpha_n}$ as the expectation of $x_1^{\alpha_1} x_2^{\alpha_2},\dots,x_n^{\alpha_n}$ w.r.t. the charge μ. Note that, $y_0 = L(1) = 1$ implies that $\int 1 d\mu = 1$ under this interpretation (normalization). Therefore, we can interpret $M_{n,d}(y)$ as a truncated moment matrix. However, since C does not include all nonnegative gambles, we cannot conclude that the charges are non-negative or, in other words, that μ is a probability charge. The constraint $M_{n,d}(y) \geq 0$ is not strong enough to guarantee non-negativity of μ (it is only a necessary condition). Negative probabilities are a manifestation of incoherence, that is they are a manifestation of the assumption of bounded rationality. Finally, the dual of the lower prevision problem (10) is then given by the convex SDP problem: $\inf_{y \in \mathbb{R}^{s_n(2d)}} L(h), \quad s.t. \quad L(g) \geq 0, \quad L(1) = 1, \quad M_{n,d}(y) \geq 0.$

Example 1 *Consider the case $n = 1, d = 1$. The matrix $M_{1,2}(y)$ is in this case*

$$M_{1,2}(y) = L\left(\begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \end{bmatrix}\right) = \begin{bmatrix} 1 & y_1 \\ y_1 & y_2 \end{bmatrix}.$$

Assume that $G = \{g_1, g_2\} = \{x_1 - 0.5, -x_1 + 0.5\}$ and so $L(g_1) = L(x_1 - 0.5) = y_1 - 0.5$ and $L(g_2) = L(-x_1 + 0.5) = -y_1 + 0.5$. Hence, we have that

$$C^\bullet = \left\{ [y_1, y_2]^\top \in \mathbb{R}^2 : y_1 - 0.5 \geq 0, \; -y_1 + 0.5 \geq 0, \; M_{1,2}([y_1, y_2]^\top) \geq 0 \right\}. \tag{17}$$

The first two constraints imply that $y_1 = 0.5$ and so we are left with the only constraint $\det(M_{1,2}([y_1, y_2]^\top)) = y_2 - 0.25 \geq 0$. Assume that we aim at computing $\underline{P}^*[-x_1(1 - x_1)]$. The solution of (10) is $\underline{P}^*[-x_1(1 - x_1)] = -0.25$ and it is attained for instance by the charge $0.352\delta_{0.367} + 0.786\delta_{0.521} - 0.138\delta_{0.281}$ (that is not a probability), here δ_a denoted an atomic charge (Dirac's delta) centred on a.

5.2 BADG as an approximating theory for ADG

We are going to show that we can use BADG as a computable approximating theory for ADG. So let us consider the BADG set $C = \mathsf{posi}(G \cup \Sigma_{2d})$ and the corresponding ADG set $\mathcal{K} = \mathsf{posi}(G \cup \mathcal{L}^+)$ (same G). We have the following result.

Theorem 10 *Assume that $\mathcal{K}$ avoids sure loss and let $f \in \mathcal{L}_{2d}$, then BADG is a conservative approximation of ADG theory in the sense that $\underline{P}^*(f) \leq \underline{P}(f)$.*

Proof Let λ_0^* be the supremum value of λ_0 such that $h - \lambda_0 - \sum_j^{|G|} \lambda_j g_j \in \Sigma_{2d}$ and λ_0^{**} the value such that $h - \lambda_0 - \sum_j^{|G|} \lambda_j g_j \geq 0$. Since the constraint $h - \lambda_0 - \sum_j^{|G|} \lambda_j g_j \in \Sigma_{2d}$ is more demanding than $h - \lambda_0 - \sum_j^{|G|} \lambda_j g_j \geq 0$, it follows that $\lambda_0^* \leq \lambda_0^{**}$. ■

The fact that $\underline{P}[f]$ is equal to the minimum of f when G is empty, i.e., Alice is in a state of full ignorance, explains why SOS polynomials are used in optimization, i.e., $\underline{P}^*[f]$ provides a lower bound for the minimum of f (Lasserre, 2009).

6. Updating

We assume that Alice considers an event "indicated" by a certain finite set of polynomial constraints $A = \{h_1(x) \geq 0, \ldots, h_{|A|}(x) \geq 0\}$: that means that Alice knows that x belongs to the set $A = \{x \in \mathbb{R}^n : h_1(x) \geq 0, \ldots, h_{|A|}(x) \geq 0\}$. In ADG we will use this information to update (conditioning) her set of desirable gambles based on A (Walley, 1991; Couso and Moral, 2011) : $\mathcal{K}_{|A} = \{g \in \mathcal{L} : gI_A \in \mathcal{K}\}$, where I_A is the indicator function on A. How do we do that in the BADG framework? In BADG we cannot completely use this information because Σ_{2d} does not include indicator functions. However, we can still exploit the information in A in a weaker way. In fact, if we know that $h_i(x) \geq 0$ in A, then we also know that $\sigma_1(x)h_1(x) + \cdots + \sigma_{|A|}(x)h_{|A|}(x) \geq 0$ $\forall x \in A$ and for every $\sigma_i \in \Sigma_{2d}$ with degree equal to $d - \lceil n_{h_i}/2 \rceil$, where n_{h_i} is the degree of $h_i(x)$ (so that the degree is less than $2d$).

Definition 11 *Let G be a finite subset of $\mathcal{L}_{2d}$, and $C = \mathsf{posi}(G \cup \Sigma_{2d})$ be a set of BADG. Assume A is a finite set of polynomial constraints. Then, the set $C_{|A}$ that includes all the gambles $f \in \mathcal{L}_{2d}$ such that there exist $\lambda_i \geq 0$, with $i = 1, \ldots, |G|$, and $\sigma_0, \sigma_1, \ldots \sigma_{|A|}, \sigma_{|A|+1} \in \Sigma_{2d}$:*

$$f - \sum_{i=1}^{|G|} \lambda_i g_i = \sigma_0 + \sum_{i=1}^{|A|} \sigma_i h_i \quad \text{and} \quad - \sum_{i=1}^{|G|} \lambda_i g_i = \sigma_{|A|+1} \tag{18}$$

is called the **updated set of desirable gambles** *based on A.*

In the state of full ignorance, since G is empty, there is only one constraint $f = \sigma_0 + \sum_{i=1}^{|A|} \sigma_i h_i$.

Theorem 12 *Let G be a finite subset of $\mathcal{L}_{2d}$, and A be a finite set of polynomial constraints. Assume that $\mathcal{K} = \mathsf{posi}(G \cup \mathcal{L}^+)$ avoids sure loss and let $f \in \mathcal{L}_{2d}$. Then we have that $\underline{P}_{C_{|A}}(f) \leq \underline{P}_{\mathcal{K}_{|A}}(f)$ where $C = \mathsf{posi}(G \cup \Sigma_{2d})$.*

Proof From the definition of conditioning for ADG we aim to find the supremum λ_0 such that $(f - \lambda_0)I_A - \sum_{j=1}^{|G|} \lambda_j g_j(x) \geq 0 \quad \forall x \in \mathbb{R}^n$. It can be rewritten as the two constraints on the left and relaxed to the constraints on the right:

$$\begin{array}{ll} - \sum_{j=1}^{|G|} \lambda_j g_j(x) \geq 0 \;\forall x \notin A, & - \sum_{j=1}^{|G|} \lambda_j g_j(x) = \sigma_{|A|+1} \;\forall x \in \mathbb{R}^n, \\ f - \lambda_0 - \sum_{j=1}^{|G|} \lambda_j g_j(x) \geq 0 \;\forall x \in A, & f - \lambda_0 - \sum_{j=1}^{|G|} \lambda_j g_j(x) = \sigma_0 + \sum_{i=1}^{|A|} \sigma_i h_i \;\forall x \in \mathbb{R}^n. \end{array}$$

■

Corollary 13 *The dual of $C_{|A}$ is*

$$C^{\bullet}_{|A} = \left\{ y \in \mathbb{R}^{s_n(d)} : L(g) \geq 0, \;\; M_{n,d}(y) \geq 0, \;\; M_{n,d-\lceil n_h/2 \rceil}(hy) \geq 0 \; \forall g \in C_{|A} \right\}$$

where $M_{n,d-\lceil n_h/2 \rceil}(hy) = L(h(x) v_{d-\lceil n_h/2 \rceil}(x) v_{d-\lceil n_h/2 \rceil}(x)^{\top})$ is called localizing matrix (Lasserre, 2009).

7. Numerical example

Consider the case $n = 2, d = 3$ and assume that Alice finds these gambles to be desirables

$$G = \{g_1, \ldots, g_7\} = \{-x_1^4 x_2^2 - x_1^2 x_2^4 + x_1^2 x_2^2 - 1, x_1, 1 - x_1, x_2, 1 - x_2, 10 - x_1^2, 10 - x_2^2\}$$

Alice first checks if her set of desirable gambles satisfies A.2 by solving (11). The solution is $\lambda_0^* = 0.0062$ and, therefore, since $\lambda_0^* \approx 0$ Alice may think that G does not incur in sure loss. To numerically verify this statement, she can increase the degree d. For $d = 4$, Alice gets $\lambda_0^* = 0.0774$ that is greater than previous solution and for $d = 5$ $\lambda_0^* = 1$. Therefore, this shows that G actually incurs a sure loss. In this case, since $\arg\max_{i>0} \lambda_i^* = 1$, the polynomial that contributes more to the sure loss is g_1.

Alice can verify if g_1 is negative by computing the BADG lower prevision of $-g_1$ for an empty G (this gives the minimum of $-g_1$ in ADG). The solution of (10) is $\underline{P}^*[-g_1] = -5.056$ for $d = 3$. For $d = 4$ we obtain $\underline{P}^*[-g_1] = 0.596$, for $d \geq 7$ $\underline{P}^*[-g_1] = 0.963$. Therefore, g_1 is strictly negative. It can be verified that 0.963 is the minimum of $-g_1$ and, therefore, $\underline{P}^*[-g_1] = \underline{P}[-g_1]$. So we have generated a family of BADG approximations (relaxations of coherence) that converge to ADG. Why can BADG obtain a lower "lower prevision" than ADG? In ADG $\underline{P}[-g_1]$ is attained by an atomic charge on the values of x_1, x_2 corresponding to the minimum of $-g_1$. Conversely, in BADG, since we allow mixtures of atomic charges with possibly negative weights, then we have more freedom in the minimization.

Alice can then remove g_1 from G and check if the following set satisfies A.2: $G \backslash g_1 = \{g_2, \ldots, g_7\} = \{x_1, 1 - x_1, x_2, 1 - x_2, 10 - x_1^2, 10 - x_2^2\}$. To prove that, Alice can solve the linear programming problem (12) that gives the solution $\lambda_0^* \approx 10^{-17}$ and shows that G avoids sure loss.

Let $f = x_1^4 + 4x_1^3 + 5.375x_1^2 + 2.75x_1 + 0.41016$ and assume Alice aims to solve (10), i.e., to compute the BADG lower prevision of f. The result is $\underline{P}^*[f] = 0.41016$ for $d \geq 3$. Now let us assume $h(x_1) = 0.0025 - (x_1 + 0.425)^2$ and compute the conditional lower prevision. The solution is $\underline{P}^*[f|A] = -0.0625$ that gives the conditional lower prevision for BADG. This is also the minimum of f in $h(x) > 0$ and coincides with the conditional lower prevision for ADG $\underline{P}[f|A]$.

8. Conclusions

In this paper we have presented a computable theory of desirable gambles by imposing bounded rationality. To achieve that we have exploited recent results from Sum-Of-Square (SOS) polynomials optimization. As future work, we plan to further develop this theory by introducing other probabilistic operations such as marginalisation and structural judgements such as epistemic independence.

References

C. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2007.

A. Benavoli and D. Piga. A probabilistic interpretation of set-membership filtering: Application to polynomial systems through polytopic bounding. *Automatica*, 70:158 – 172, 2016.

I. Couso and S. Moral. Sets of desirable gambles: Conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52:1034–1055, 2011.

B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937.

D. Hilbert. Über die darstellung definiter formen als summe von formenquadraten. *Mathematische Annalen*, 32(3):342–350, 1888.

J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2009.

E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008. doi:10.1016/j.ijar.2007.12.001.

E. Miranda and M. Zaffalon. Notes on desirability and conditional lower previsions. *Annals of Mathematics and Artificial Intelligence*, 60(3–4):251–309, 2010.

R. Pelessoni and P. Vicig. 2-coherent and 2-convex conditional lower previsions. *International Journal of Approximate Reasoning*, 77:66–86, 2016.

D. Piga and A. Benavoli. A unified framework for deterministic and probabilistic *D*-stability analysis of uncertain polynomial matrices. *IEEE Transactions on Automatic Control, In press*, 2018.

M. J. Schervish, T. Seidenfeld, and J. B. Kadane. How sets of coherent probabilities may serve as models for degrees of incoherence. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(03):347–355, 2000.

A. Seidenberg. A new decision method for elementary algebra. *Annals of Mathematics*, pages 365–374, 1954.

T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Decisions without ordering. In W. Sieg, editor, *Acting and reflecting*, volume 211 of *Synthese Library*, pages 143–170. Kluwer, Dordrecht, 1990.

A. Tarski. A decision method for elementary algebra and geometry. 1951.

M. Troffaes and G. De Cooman. Extension of coherent lower previsions to unbounded random variables. *Intelligent systems for information processing: from representation to applications*, pages 277–288, 2003.

P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

P. Walley, R. Pelessoni, and P. Vicig. Direct algorithms for checking coherence and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, pages 119–151.

P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.